(12) UK Patent Application (19) GB (11) 2 326 781 (13) A

(43) Date of A Publication 30.12.1998

(21) Application No 9711313.8

(22) Date of Filing 30.05.1997

(71) Applicant(s)
British Broadcasting Corporation
(Incorporated in the United Kingdom)
Broadcasting House, LONDON, W1A 1AA,
United Kingdom

(72) Inventor(s)
Nicholas Dominic Wells

(74) Agent and/or Address for Service
Mathys & Squire
100 Grays Inn Road, LONDON, WC1X 8AL,
United Kingdom

(51) INT CL[6]
H04N 5/04 , G11B 27/10

(52) UK CL (Edition P)
H3A ARX
H4F FBB FD3P FD3T FD30K FD32
U1S S2204

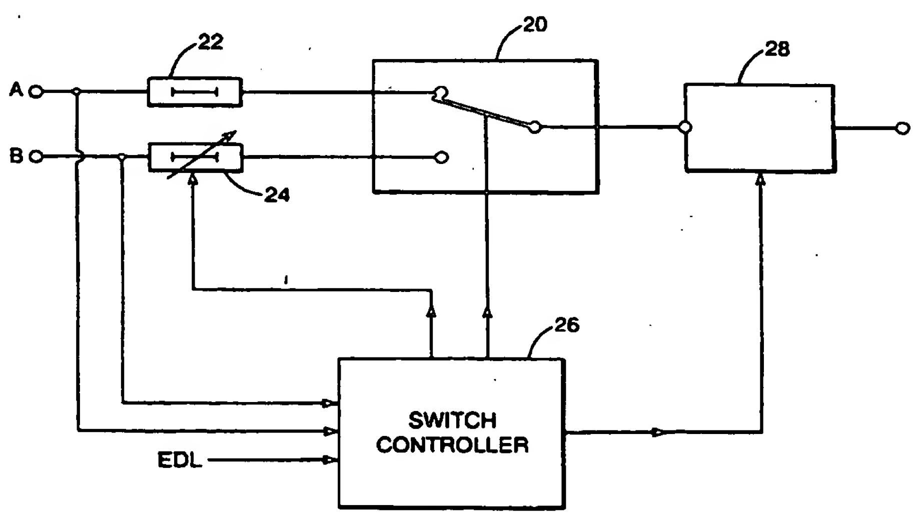
(56) Documents Cited
US 5570372 A
INSPEC abstract of non-patent literature:"Maintaining audio and video synchronisation in television" by Kirby et al in Proceedings of 1994 International Broadcasting Convention- IBC'94, p.121-6, IEE, London, UK, 1994.

(58) Field of Search
UK CL (Edition O) G5R RB81 , H3A ARC ARD ARX , H4F FBB
INT CL[6] G11B , H04H , H04N
Online: WPI, JAPIO, INSPEC

(54) Abstract Title
Video-audio synchronization

(57) In the processing of digital television signals, to maintain lip-sync when a video signal has been edited or switched at a video frame boundary, the associated audio signal is advanced or retarded to the closest audio frame boundary. Any error introduced by this constraint is accumulated and carried forward to direct a subsequent advance or retard decision. In this way, the cumulative error can be kept within an acceptable tolerance. In the arrangement shown, an audio switch 20 receives an audio signal A through fixed delay 22 and an audio signal B through a variable delay 24. A switch controller 26 receives an Edit Decision List prescribing the desired point of switching from A to B. The frame timings are determined from signal A and the current accumulated sub-frame synchronization error is determined from signal B. The controller decides whether to advance or retard signal B. A residual sub-frame synchronization is calculated and optionally embedded at 28.

FIG. 2



At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

GB 2 326 781 A

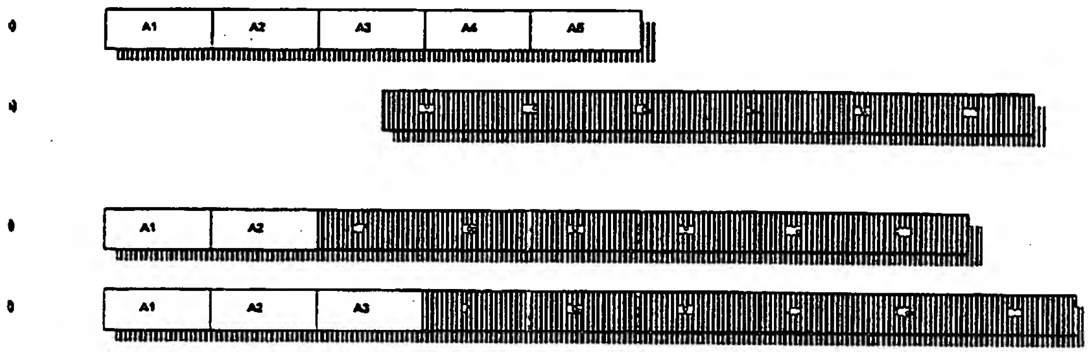


Figure 1: Examples of switching audio at frame boundaries (Audio frame period = 24 ms)

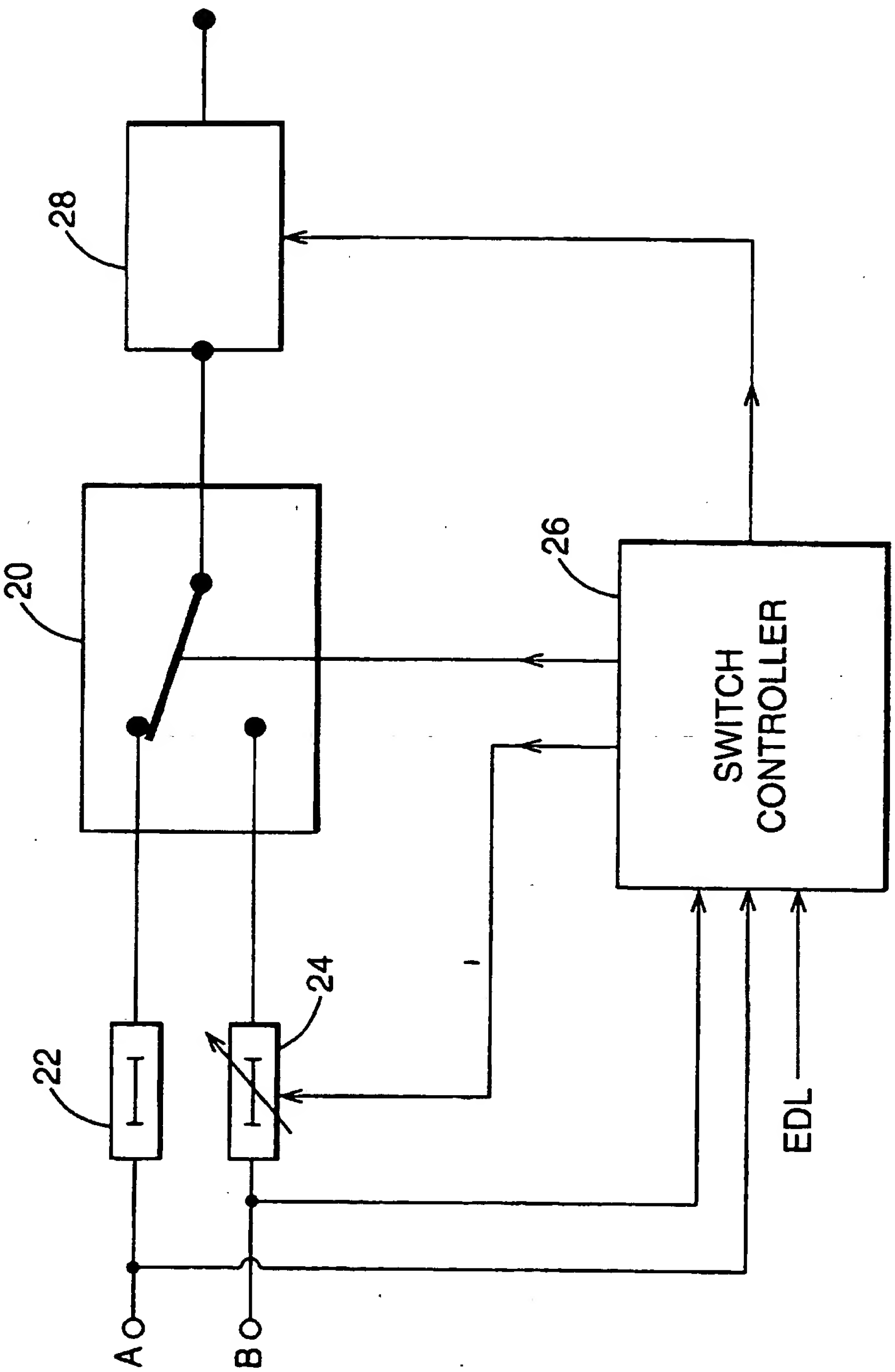
a) Bitstream A (co-timed with video A)

b) Bitstream B (co-timed with video B)

c) Switched bitstream with Audio B advanced

b) Switched bitstream with Audio B delayed

FIG. 2



VIDEO AND AUDIO SIGNAL PROCESSING

This invention relates to the editing and switching of digital television signals consisting of video and associated sound components. It is particularly relevant to systems which handle the audio component in compressed form.

The component parts of the TV signal necessarily carry with them information which enables their relative replay/display timing to be maintained. However, if for any reason, the replayed/displayed audio/video timings get out of step then this is most noticeable when people are speaking. Consequently, an error in audio/video replay/display timings is often referred to as "lip-sync error".

A variety of techniques have been proposed for maintaining lip-sync in circumstances where the video and audio components have undergone separate processing with the introduction of different delays. Reference is directed, for example, to GB-B-2 273 215 which discloses a technique by which a delay signal is associated with either the video or audio signal, that delay signal representing the relative delay between video and audio. Signal processing elements which introduce delay are provided with synchronizing units which read the delay signal at the input, add or subtract the appropriate delay interval and insert the modified delay signal at the output. At the end of a signal processing chain, the delay signal is read and the appropriate delay interval is inserted. Since there is usually more signal processing conducted on the video component, the greater delay will usually be suffered by the video signal. The compensating delay at the end of the signal processing chain is typically an audio delay only. If a video delay is required, GB-B-2 273 215 explains that this can only be adjusted in coarse steps in order to maintain synchronization between its input and output. An additional audio delay will then be used to restore precise synchronization.

The present invention recognizes that in a number of key applications, the audio signal itself will be arranged in discrete frames having a defined time interval. It will no longer be possible in such applications to achieve precise synchronization – to any desired level of accuracy – through coarse

frame–by–frame control of video delay and fine, continuous control of audio delay. The prior art proposals, when audio processing is necessarily organised into discrete audio frame intervals, will enable synchronization to be achieved only to within the resolution of one audio frame. The time intervals of audio frames – as will later be discussed in more detail – are considerably shorter than video frames and a lip–sync error of a fraction of an audio frame time interval may not be objectionable. The difficulty is that the effect of such sub–frame synchronization errors is cumulative and conventional equipment may still produce objectionable lip–sync errors even if every delay within the chain has been defined within the accuracy of one audio frame.

It is an object of the present invention to provide improved methods for use in editing, switching and like signal processing of temporally associated frame–based signals, which address this problem.

It is a further object of this invention to provide improved methods for use in editing, switching and like signal processing where sub–frame synchronization errors are controlled or managed.

Accordingly, the present invention consists in one aspect in a method for use in the editing and switching of video and associated audio signals in which editing and switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries, wherein accumulated sub–frame synchronization error information is carried forward.

Preferably, accumulated sub–frame synchronization error information is employed to direct a frame advance/retard decision in a frame accurate synchronization process.

Suitably, said advance/retard decision operates on audio frames.

Advantageously, accumulated sub–frame synchronization error information is used to keep the total sub–frame synchronization error introduced by the cascading of processing stages within certain tolerances.

In another aspect, the present invention consists in a method for use in editing, switching and like signal processing of temporally associated

frame-based signals, in which steps are taken to ensure frame accurate synchronization, characterized in that sub-frame synchronization error feedback is provided to minimize cumulative sub-frame synchronization error.

In yet another aspect, the present invention consists in a method for use in editing, switching and like signal processing of temporally associated frame-based signals which are provided in digital form with information prescribing the time of presentation of frames, characterized by the maintenance through said processes of signal synchronization information; independent of said information prescribing the time of presentation of frames.

The invention will now be described by way of example with reference to the accompanying drawings, in which:-

Figure 1 is a block diagram illustrating the synchronization of frame based audio signals; and

Figure 2 is a block diagram illustrating apparatus performing a method according to one embodiment of this invention.

When digital audio is compressed, the compressed audio data is usually organized into "audio frames". For example, in MPEG2 Layer II auio coding, the audio frames are 24ms long.

When switching/editing such audio data it may be important to switch between source audio clips at times coincident with audio frame boundaries (for example in order to avoid or minimize the build up of re-quantisation noise). If this is the case then, when switching/editing video and associated audio signals it is unlikely that it will be possible to maintain continuity and regularity of video frames and simultaneously maintain continuity and regularity of audio frames without introducing some lip-sync error after the switch/edit point. This is because the video frame period and the audio frame period are not equal or related in a very simple fashion.

For the MPEG Layer II example the maximum lip-sync error introduced at a switch/edit point could be up to +/- 12ms as shown in Figure 1.

Along the process of programme production, continuity switching and distribution, the signal may be switched and edited several times, and successive errors could accumulate to an unacceptable value. Therefore, some mechanism is required to prevent the build up of lip-sync errors through successive generations of editing/switching. This mechanism is in addition to other mechanisms which may be employed to ensure the overall audio/video synchronization at points along the combined signal chain.

Coded digital audio frames usually contain bytes for signalling framing information together with information as to how to decode that audio frame. Also, there are usually slots in the audio frame which can be used to carry user-specified data.

The proposed solution to the lip-sync error problem is to use some of these user-specified data bytes to signal the lip-sync error introduced by an edit switch process. At each edit switch process in the production and distribution chain, any equipment performing a switch can read the existing value in the lip-sync field and decide whether to delay or advance the switched audio (by including or removing one extra audio frame) such that the overall lip-sync error is minimized after the new switch edit point. Following the switch point the equipment should re-insert the new value for the total lip-sync error in the lip-sync error field.

For the example of MPEG Layer II switching and editing, this technique would enable the total lip-sync error to be kept within +/- 12 ms. Some systems may be defined which switch on sub-frame boundaries (e.g. at 8ms intervals), but this technique would still be important to prevent unacceptable accumulation of lip-sync errors.

It should be noted that compressed bitstreams such as those defined by the MPEG standard contain information, called *Presentation Time Stamps (PTS),* which instruct a decoder at what time a given audio or video frame should be replayed/displayed. It is intended that this *PTS* information is the means by which the audio and video should be replayed/displayed in synchronism at the output of a decoder. However, in situations such as that described by Figure 1, the decoder is being instructed by the audio *PTS*

values to decode the audio with a possible lip-sync error. It is not possible to simply change the audio *PTS* values for the audio frames following a switch point because this would lead to a discontinuity in the decoded audio signal.

For audio (and video) editing systems, the audio/video signal is replayed from disk according to an *edit list* or a *playout list.* In this case, the audio/video replay/display timing of any particular audio/video frame is specified through the *edit list* rather than via *PTS* values. In switching between separate audio source clips, the requirement for continuity of audio frames will inevitably lead to lip-sync errors as shown in Figure 1.

In summary therefore, editing/switching equipment should perform the following operations around each edit/switch point:

i) Note the time at which an audio frame should be replayed in order to maintain synchronism with the associated video. (This information might come from synchronization information carried in the bitstream or from an edit list).

ii) Read the lip-sync-error field.

iii) In order to maintain continuity of audio frames (or some sub-multiple of the audio frame period), decide whether to advance or delay the following sequence of audio frames in a manner which minimizes the accumulated lip-sync error.

iv) Re-insert in the each frame of the audio a new value corresponding the new accumulated lip-sync error.

Referring now to Figure 2, an audio switch 20 receives audio signal A through fixed delay 22 and audio signal B through variable delay 24. A switch controller 26 receives an Edit Decision List (EDL) prescribing in this case the desired point of switching from A to B. The switch controller additionally receives audio signal A, from which it determines the frame timings, and audio signal B, from which it determines the current accumulated sub-frame synchronization error. Utilising this information, the switch controller decides whether to advance or retard signal B at the switch point, as illustrated in Figure 1. To enable signal B to be effectively

advanced or retarded, a fixed delay is inserted in the path of signal A and the switch controller selects the appropriate value for the variable delay in the path of signal B. A residual sub-frame synchronization error is calculated and provided as an output for optional embedding in the switched output through embedder 28.

It will be recognised that for simplicity, apparatus has been shown which is capable of switching from signal A to signal B only; a symmetrical version of the described apparatus would also permit switching from signal B to signal A.

In one project, a TV switching and editing system is being developed which uses compressed MPEG2 coded video and MPEG Layer II coded audio. The audio is stored in the form of Packetised Elementary Streams (PES) with one audio frame per PES packet. The PES packet consists of a header of 40 bytes followed by the bytes of the compressed audio frame data. The header contains 16 bytes of user data. Two bytes of this user data field have been reserved to signal the lip-sync error in units of 90kHz clock periods.

In some applications switching may be done using audio signals which have been decoded from previously compressed data. The switched signal may then be re-compressed and/or passed on to equipment which can correct for any lip-sync error. In these cases, it is beneficial to be able to carry lip-sync error information within the decoded audio signal.

Several methods for carrying such data within a decoded audio signal which may be in digital or analogue form have been described in a previous patent application [No. GB 9701616.6].

For decoded audio signals, the concept of *audio frame* can still be applied. In the decoded domain, an audio frame has a one-to-one correspondence with the relevant audio frame in the compressed domain.

The ideas described in this document can also be applied to the switching and editing of any program related data (PRD) where relative timing of the replay of the PRD signal needs to be maintained to within a given tolerance in relation to other signals in the digital multiplex.

CLAIMS

1. A method for use in the editing and switching of video and associated audio signals in which editing and switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries, wherein accumulated sub-frame synchronization error information is carried forward.

2. A method according to Claim 1, in which accumulated sub-frame synchronization error information is employed to direct a frame advance/retard decision in a frame accurate synchronization process.

3. A method according to Claim 2, in which said advance/retard decision operates on audio frames.

4. A method according to any one of the preceding claims, in which accumulated sub-frame synchronization error information is carried as data within compressed audio data; for example in the digital header of a compressed audio frame.

5. A method according to any one of Claims 1 to 3, in which accumulated sub-frame synchronization error information is carried as data within a decoded/uncompressed digital or audio signal.

6. A method according to any one of the preceding claims, in which accumulated sub-frame synchronization error information is used to keep the total sub-frame synchronization error introduced by the cascading of processing stages within certain tolerances.

7. A method for use in editing, switching and like signal processing of temporally associated frame-based signals, in which steps are taken to ensure frame accurate synchronization, characterized in that sub-frame

synchronization error feedback is provided to minimize cumulative sub-frame synchronization error.

8. A method for use in editing, switching and like signal processing of temporally associated frame-based signals which are provided in digital form with information prescribing the time of presentation of frames, characterized by the maintenance through said processes of signal synchronization information, independent of said information prescribing the time of presentation of frames.

The Patent Office

9

Application No: GB 9711313.8
Claims searched: 1-6

Examiner: Mr.Sat Satkurunath
Date of search: 26 August 1997

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.O): H3A: ARC, ARD, ARX; H4F: FBB; G5R: RB81

Int Cl (Ed.6): G11B, H04H, H04N

Other: Online: WPI, JAPIO, INSPEC

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	US 5570372 SIEMENS - see especially figure 1 and abstract	1
A	INSPEC abstract of non-patent literature "Maintaining audio and video synchronisation in television" by Kirby et al in Proceedings of 1994 International Broadcasting Convention - IBC`94, P121-6, IEE, London, UK, 1994.	1

X Document indicating lack of novelty or inventive step
Y Document indicating lack of inventive step if combined with one or more other documents of same category.
& Member of the same patent family
A Document indicating technological background and/or state of the art.
P Document published on or after the declared priority date but before the filing date of this invention.
E Patent document published on or after, but with priority date earlier than, the filing date of this application.

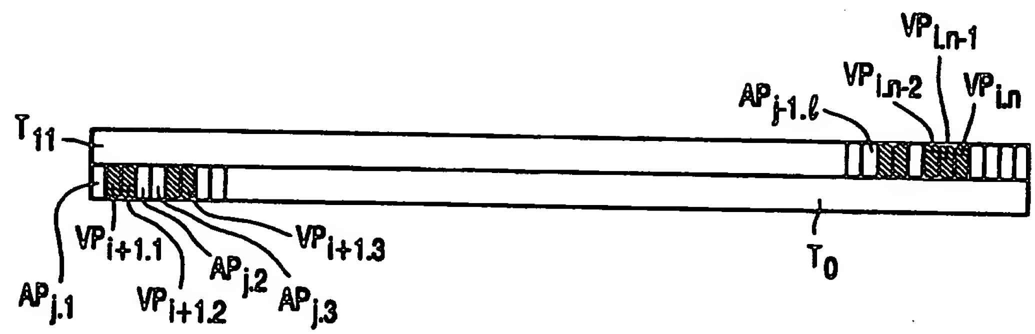


FIG. 4

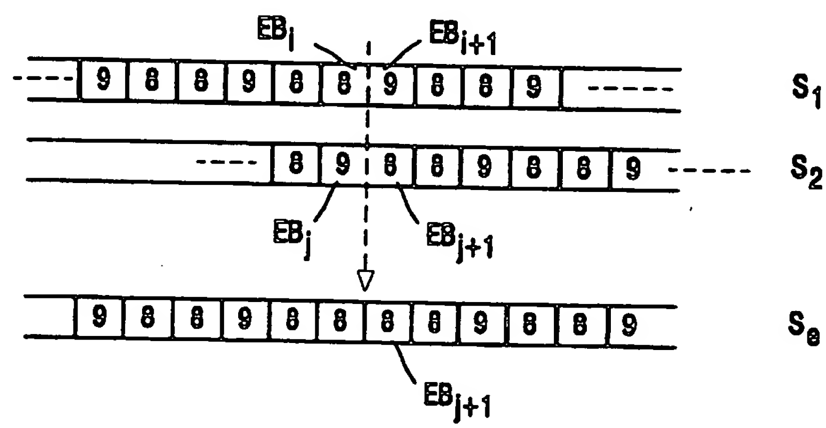


FIG. 5

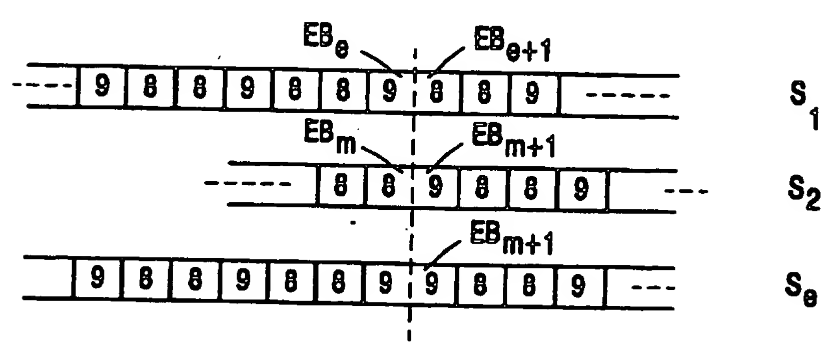


FIG. 6

3/4

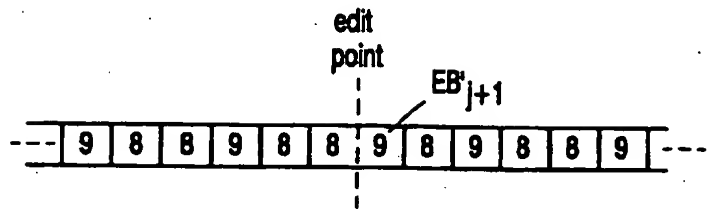


FIG. 7

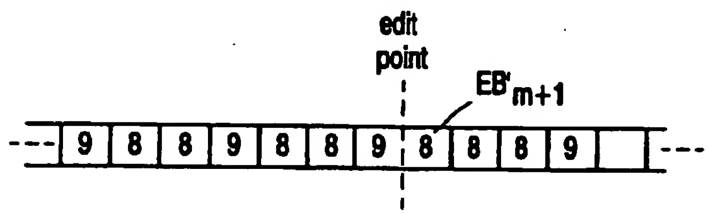


FIG. 8

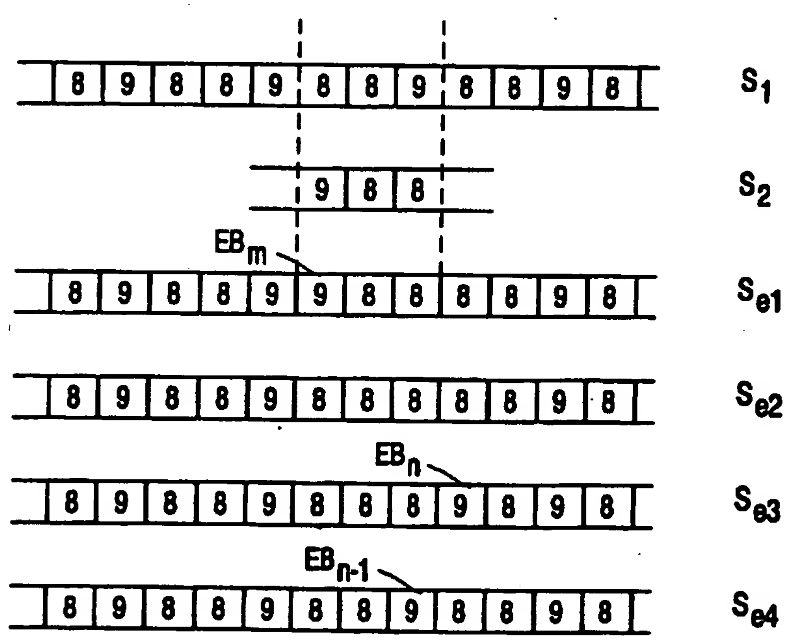
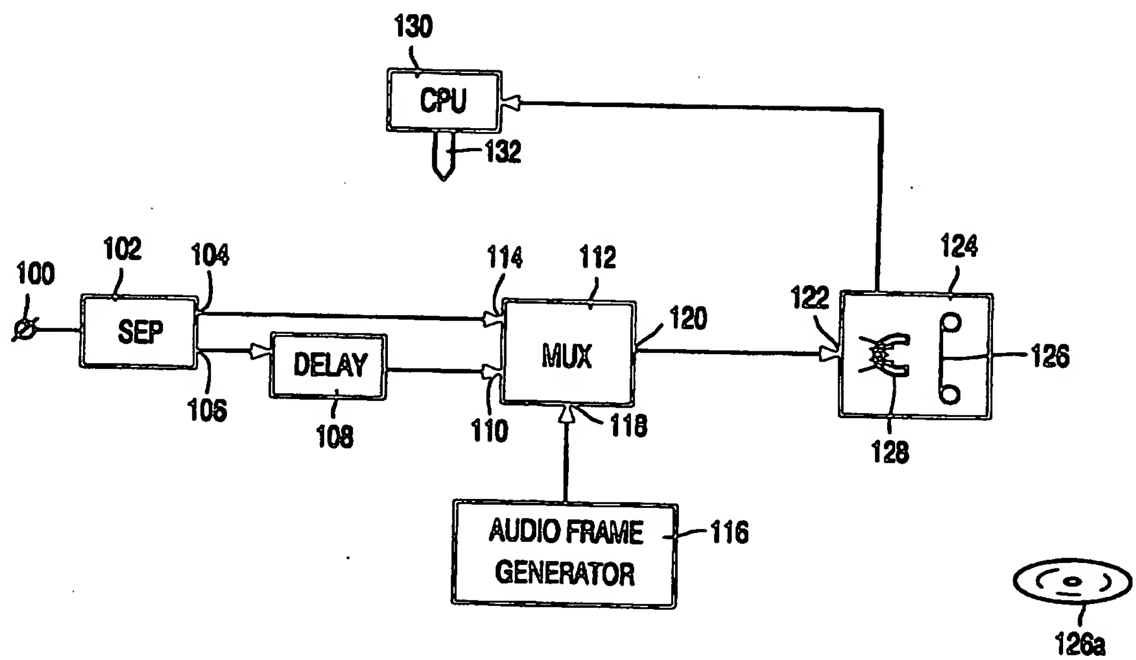


FIG. 9



130
CPU
132
100
102
104
SEP
106
DELAY
108
114
112
MUX
110
118
120
122
124
126
128
AUDIO FRAME GENERATOR
116
126a

FIG. 10